What is claimed is:

1. A method for use in sequence data analysis comprising:

providing a multiple sequence alignment of a plurality of sequences, wherein the multiple sequence alignment comprises a column of aligned amino acids and/or gaps for each horizontal position of the multiple sequence alignment;

providing a plurality of numerical physical-chemical property (PCP) descriptors for each amino acid based on a plurality of physical-chemical properties thereof, wherein each of the plurality of numerical PCP descriptors corresponds to one of "N" eigenvectors used in defining the amino acids in terms of physical-chemical properties;

describing each amino acid in the multiple sequence alignment quantitatively in terms of the plurality of PCP descriptors as a series of "N" eigenvectors resulting in "N" PCP described sequence alignments, wherein each PCP described sequence alignment corresponds to and is defined with numerical PCP descriptors which correspond to one of the "N" eigenvectors, and further wherein each PCP described sequence alignment comprises a plurality of columns corresponding to the columns of the multiple sequence alignment;

analyzing each of the PCP described sequence alignments, on a column by column basis, to generate conservation property data for each column, wherein the conservation property data for each column comprises an average value for the numerical PCP descriptors in the column and a standard deviation associated with the average value, and a relative entropy value for the column;

analyzing the conservation property data for each of the PCP described sequence alignments to detect consecutive horizontal positions of the multiple sequence alignment where the physical-chemical properties are conserved based on at least the relative entropy determined for each column; and

defining one or more PCP motifs in the multiple sequence alignment based at least on the detection of consecutive horizontal positions of the multiple sequence alignment where the physical-chemical properties are conserved according to at least one eigenvector.

2. The method of claim 1, wherein analyzing the conservation property data for each of the PCP described sequence alignments comprises analyzing the conservation property data

for each of the PCP described sequence alignments to detect consecutive horizontal positions where the relative entropy satisfies a predetermined limit.

3. The method of claim 1, wherein defining one or more PCP motifs in the multiple sequence alignment further comprises using user specified gap and minimum length limits to define the one or more PCP motifs, wherein each PCP motif comprises a plurality of consecutive horizontal positions in the multiple sequence alignment.

4. The method of claim 1, further wherein the method comprises using the one or more PCP motifs to search a sequence database for related sequence segments having PCP characteristics similar to one or more of the PCP motifs.

5. The method of claim 4, wherein each PCP motif comprises a plurality of consecutive horizontal positions in the multiple sequence alignment, wherein using the one or more PCP motifs to search a sequence database for related sequence segments comprises defining each of the PCP motifs as a series of PCP motif profile matrices, wherein each PCP motif profile matrix of the series corresponds to one of the "N" eigenvectors, and further wherein values for each PCP motif profile matrix comprise an average value of the numerical PCP descriptors in the column at each horizontal position of the PCP motif and a standard deviation associated with the average value, and a relative entropy value for each horizontal position of the PCP motif.

6. The method of claim 5, wherein using the one or more PCP motifs to search a sequence database for related sequence segments comprises:

converting each of one or more sequences of the sequence database to a searchable form using the numerical PCP descriptors;

using a positional scoring function to match values of the series of PCP motif profile matrices defined for each PCP motif to segments of each of the searchable matrices resulting in scored segments; and

selecting at least one scored segment for each of the searchable matrices as being a best match to each PCP motif based on results of the positional scoring function.

7. The method of claim 6, wherein each of the selected scored segments forms a part of one of a plurality of proteins of the sequence database, and wherein the method further comprises ranking the plurality of proteins according to which protein has PCP characteristics that are the closest to the plurality of sequences used to provide the multiple sequence alignment.

8. The method of claim 7, wherein ranking the plurality of proteins comprises ranking one or more of the plurality of proteins based on application of a Bayesian scoring function.

9. The method of claim 7, wherein ranking the plurality of proteins further comprises ranking one or more of the plurality of proteins based on structural similarity.

10. The method of claim 7, wherein ranking the plurality of proteins comprises:

determining an overall PCP similarity distance score associated with each of the one or more proteins of the sequence database; and

ranking the one or more proteins of the sequence database based on the overall PCP similarity scores for the proteins and relative to what a random score for the proteins would be.

11. The method of claim 6, wherein each of the selected scored segments forms a part of one of a plurality of proteins of the sequence database, and wherein the method further comprises:

providing structural data for the one or more selected sequence segments;

providing query structural data related to the PCP motifs;

calculating segmental root mean square deviation between the query structural data and the structural data for the one or more selected sequence segments; and

ranking the one or more proteins of the sequence database based on the calculated segmental root mean square deviation.

12. A computer program for use in conjunction with a processing apparatus to analyze

sequence data, wherein the computer program is operable when used with the processing apparatus to:

recognize a multiple sequence alignment of a plurality of sequences, wherein the multiple sequence alignment comprises a column of aligned amino acids and/or gaps for each horizontal position of the multiple sequence alignment;

recognize a plurality of numerical physical-chemical property (PCP) descriptors for each amino acid based on a plurality of physical-chemical properties thereof, wherein each of the plurality of numerical PCP descriptors corresponds to one of "N" eigenvectors used in defining the amino acids in terms of physical-chemical properties;

describe each amino acid in the multiple sequence alignment quantitatively in terms of the plurality of PCP descriptors as a series of "N" eigenvectors resulting in "N" PCP described sequence alignments, wherein each PCP described sequence alignment corresponds to and is defined with numerical PCP descriptors which correspond to one of the "N" eigenvectors, and further wherein each PCP described sequence alignment comprises a plurality of columns corresponding to the columns of the multiple sequence alignment;

analyze each of the PCP described sequence alignments, on a column by column basis, to generate conservation property data for each column, wherein the conservation property data for each column comprises an average value for the numerical PCP descriptors in the column and a standard deviation associated with the average value, and a relative entropy value for the column;

analyze the conservation property data for each of the PCP described sequence alignments to detect consecutive horizontal positions of the multiple sequence alignment where the physical-chemical properties are conserved based on at least the relative entropy determined for each column; and

define one or more PCP motifs in the multiple sequence alignment based at least on the detection of consecutive horizontal positions of the multiple sequence alignment where the physical-chemical properties are conserved according to at least one eigenvector.

13. The computer program of claim 12, wherein the computer program is operable when used with the processing apparatus to analyze the conservation property data for each of the PCP described sequence alignments by analyzing the conservation property data for each of

the PCP described sequence alignments to detect consecutive horizontal positions where the relative entropy satisfies a predetermined limit.

14. The computer program of claim 12, wherein the computer program is operable when used with the processing apparatus to define one or more PCP motifs in the multiple sequence alignment using user specified gap and minimum length limits to define the one or more PCP motifs, wherein each PCP motif comprises a plurality of consecutive horizontal positions in the multiple sequence alignment.

15. The computer program of claim 12, wherein the computer program is further operable when used with the processing apparatus to use the one or more PCP motifs to search a sequence database for related sequence segments having PCP characteristics similar to one or more of the PCP motifs.

16. The computer program of claim 15, wherein each PCP motif comprises a plurality of consecutive horizontal positions in the multiple sequence alignment, and wherein the computer program is further operable when used with the processing apparatus to define each of the PCP motifs as a series of PCP motif profile matrices, wherein each PCP motif profile matrix of the series corresponds to one of the "N" eigenvectors, and further wherein values for each PCP motif profile matrix comprise an average value of the numerical PCP descriptors in the column at each horizontal position of the PCP motif and a standard deviation associated with the average value, and a relative entropy value for each horizontal position of the PCP motif.

17. The computer program of claim 16, wherein the computer program is further operable when used with the processing apparatus to:

convert each of one or more sequences of the sequence database to a searchable form using the numerical PCP descriptors;

use a positional scoring function to match values of the series of PCP motif profile matrices defined for each PCP motif to segments of each of the searchable matrices resulting in scored segments; and

select at least one scored segment for each of the searchable matrices as being a best match to each PCP motif based on results of the positional scoring function.

18. The computer program of claim 17, wherein each of the selected scored segments forms a part of one of a plurality of proteins of the sequence database, and wherein the computer program is further operable when used with the processing apparatus to rank one or more of the plurality of proteins according to which protein has PCP characteristics that are the closest to the plurality of sequences used to provide the multiple sequence alignment.

19. The computer program of claim 18, wherein the computer program is operable when used with the processing apparatus to rank one or more of the plurality of proteins based on application of a Bayesian scoring function.

20. The computer program of claim 18, wherein the computer program is operable when used with the processing apparatus to rank one or more of the plurality of proteins based on structural similarity.

21. The computer program of claim 18, wherein the computer program is operable when used with the processing apparatus to:

determine an overall PCP similarity distance score associated with each of the one or more proteins of the sequence database; and

rank the one or more proteins of the sequence database based on the overall PCP similarity scores for the proteins and relative to what a random score for the proteins would be.

22. The method of claim 17, wherein each of the selected scored segments forms a part of one of a plurality of proteins of the sequence database, and wherein the computer program is operable when used with the processing apparatus to:

recognize structural data for the one or more selected sequence segments;

recognize query structural data related to the PCP motifs;

calculate segmental root mean square deviation between the query structural data and the structural data for the one or more selected sequence segments; and

rank the one or more proteins of the sequence database based on the calculated segmental root mean square deviation.

5